예제 2: 예 10-5

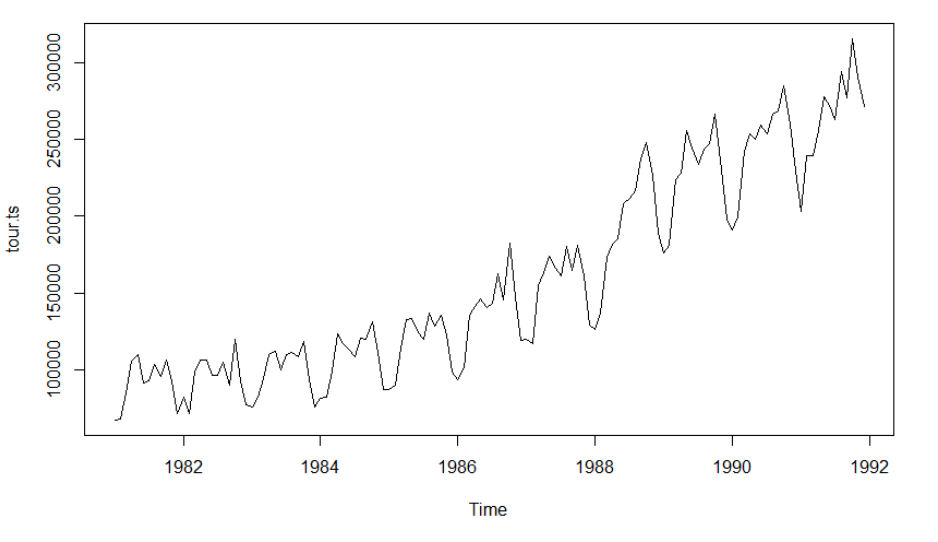
- 자료: 우리나라에 입국한 관광객 수

- 분석모형
  - 계절형 ARIMA 모형

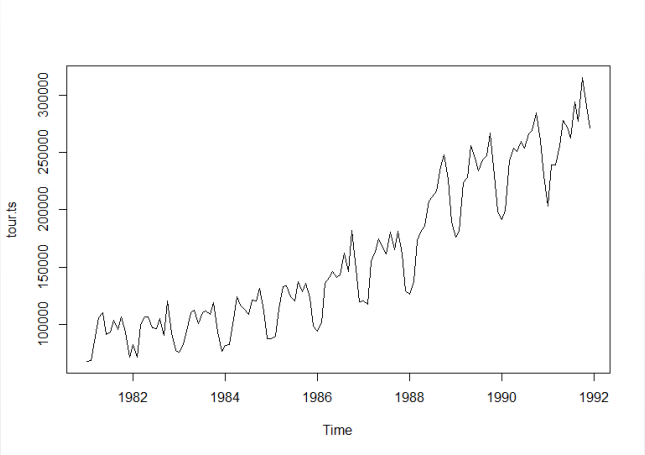
- Training data(tourist.txt): 1981.1~1991.12
- Test data(tour92.txt): 1992.1~1992.12

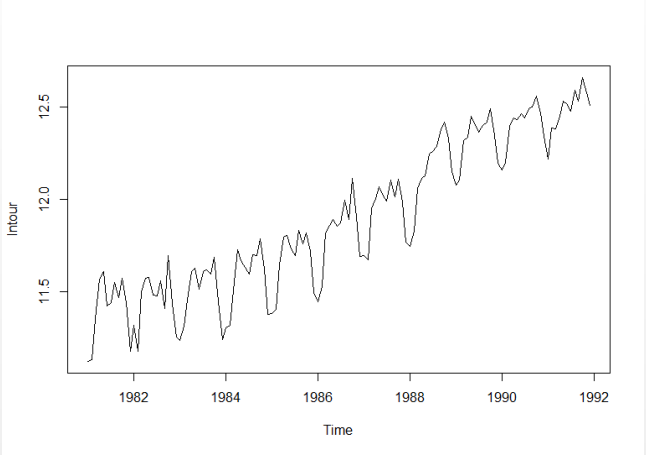
- 시계열 그림 작성

```
> tour <- scan("D:/Data/tourist.txt")
Read 132 items
> tour.ts <- ts(tour, start=1981, frequency=12)
> plot(tour.ts)
```



- 정상성 확인 1 : 동일분산 확인

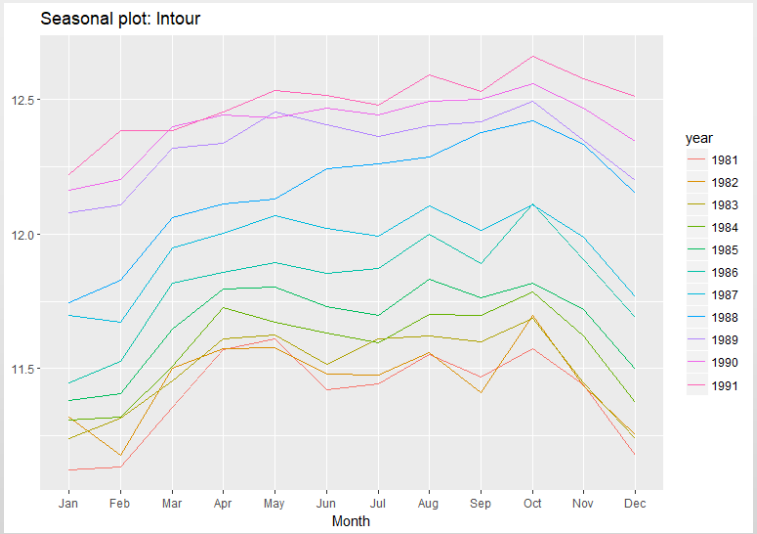




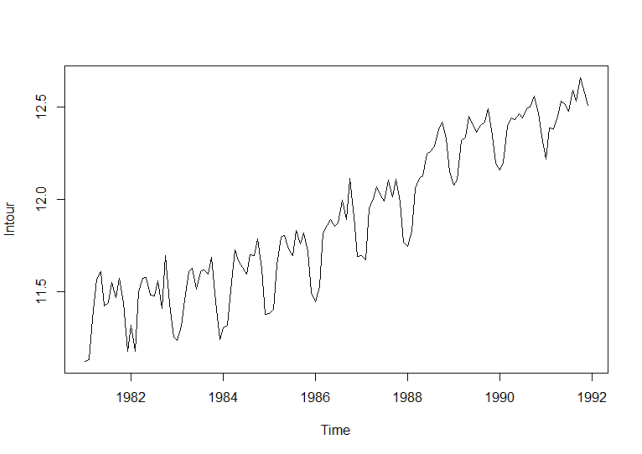
분산 안정화

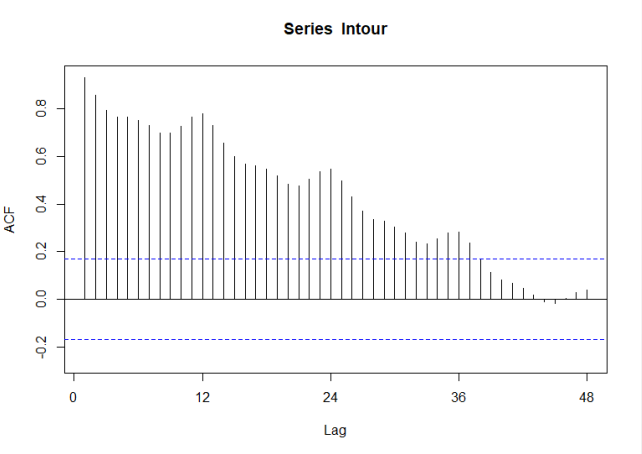
- 분산 증가
- 로그 변환으로 분산 안정화 시도

```
> ltour <- log(tour.ts)
```



- 정상성 확인 2: 추세 확인



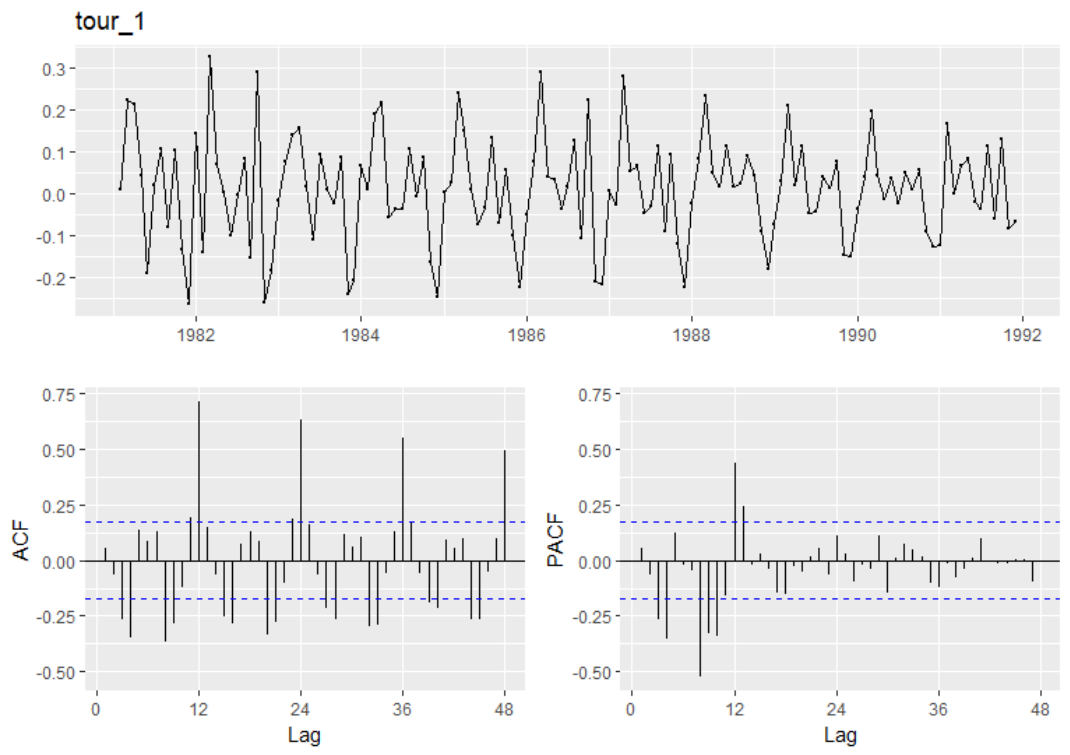



```
> library(forecast)
> ndiffs(lntour)
[1] 1
> nsdiffs(lntour,m=12)
[1] 0
```

일반 차분 and/or 계절 차분 필요

- d=1의 경우

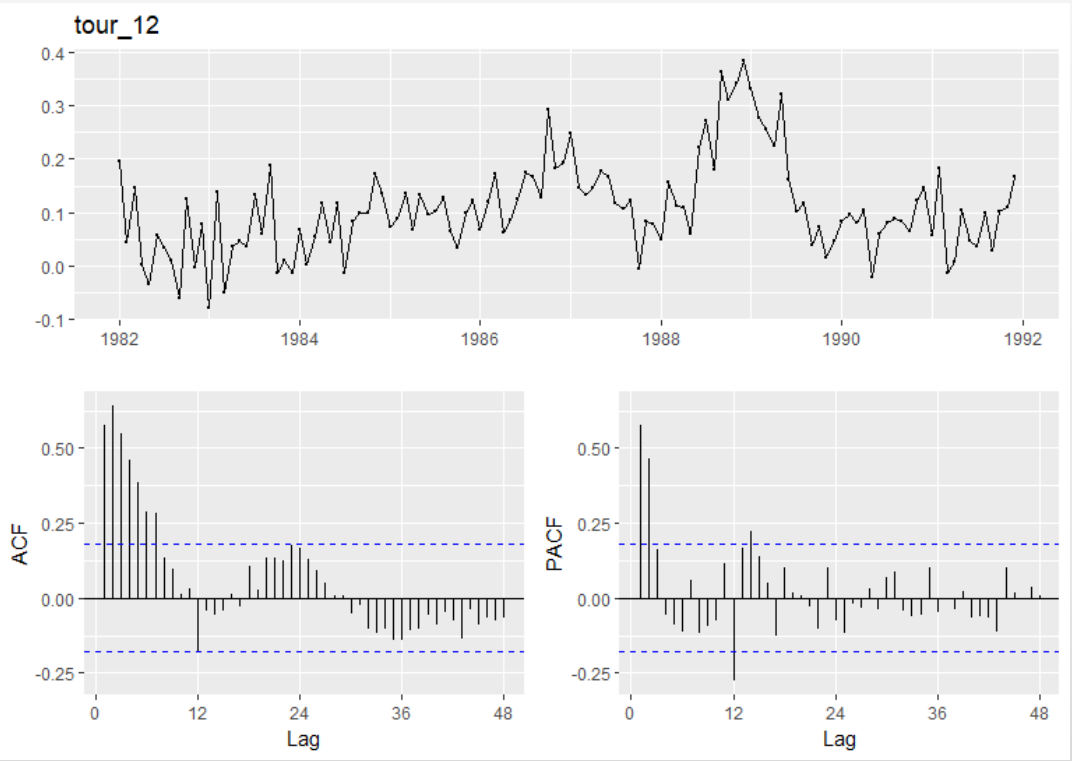
```
> tour_1 <- diff(lntour)
> ggtsdisplay(tour_1,lag.max=48)
```



추가적인
계절 차분 필요

- D=1의 경우

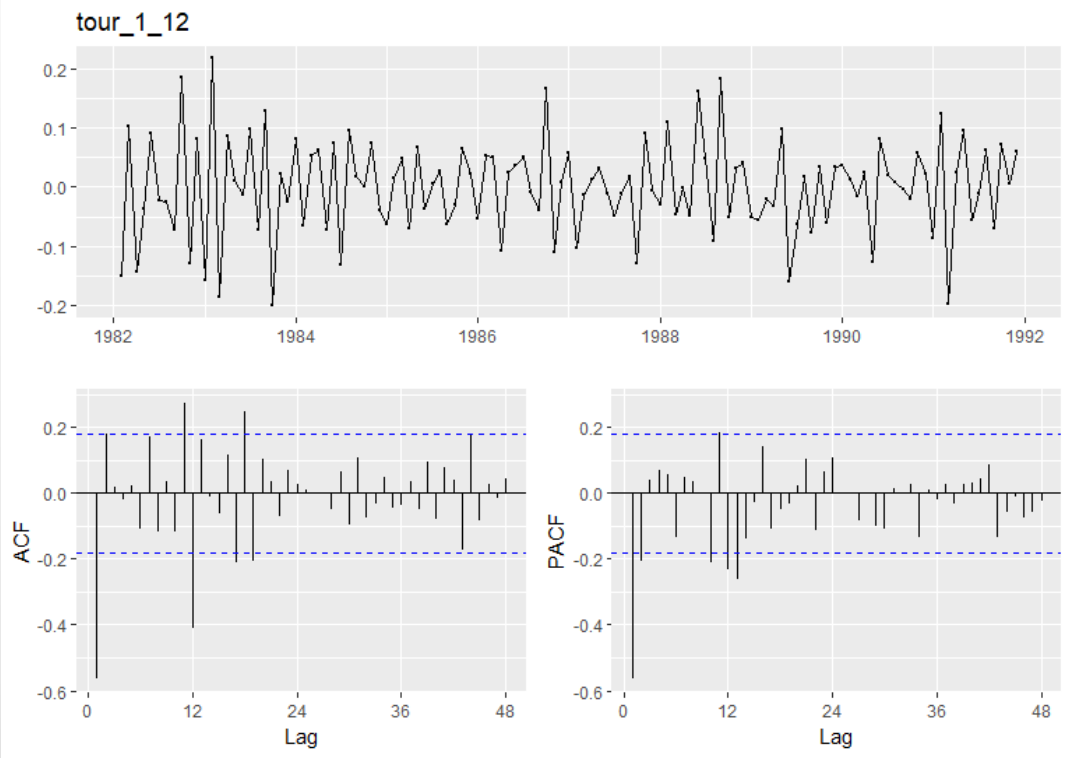
```
> tour_12 <- diff(lntour,lag=12)
> ggtsdisplay(tour_12,lag.max=48)
> ndiffs(tour_12)
[1] 1
```



추가적인
일반 차분 필요

- d=1, D=1의 경우

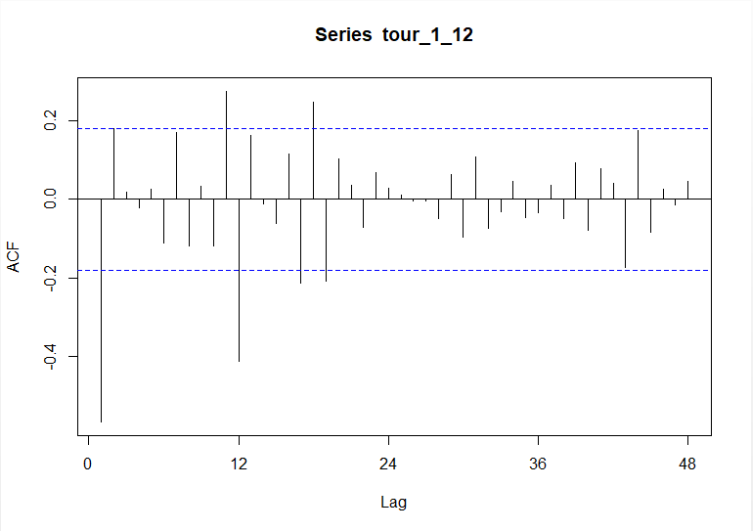
```
> tour_1_12 <- diff(tour_1,lag=12)
> ggtsdisplay(tour_1_12,lag.max=48)
```



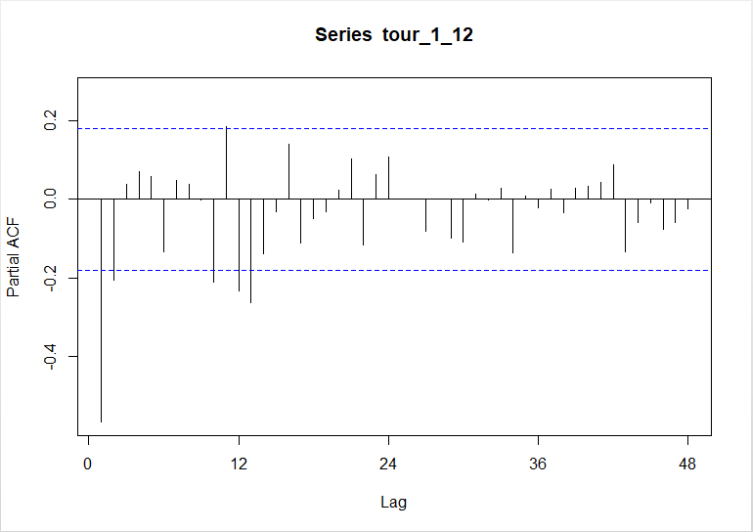
더 이상의 차분은 필요 없음

d=1, D=1으로 결정

- 모형 인식



비계절형 요소
ACF 절단, PACF 감소: (p=0, q=1)
ACF 감소, PACF 절단: (p=2, q=0)



계절형 요소
12 시차 근처에서만 유의적
→ (P=0, Q=1), (P=1,Q=0)

1) $ARIMA(0,1,1)(0,1,1)_{12}$
   $ARIMA(0,1,1)(1,1,0)_{12}$

2) $ARIMA(2,1,0)(0,1,1)_{12}$
   $ARIMA(2,1,0)(1,1,0)_{12}$

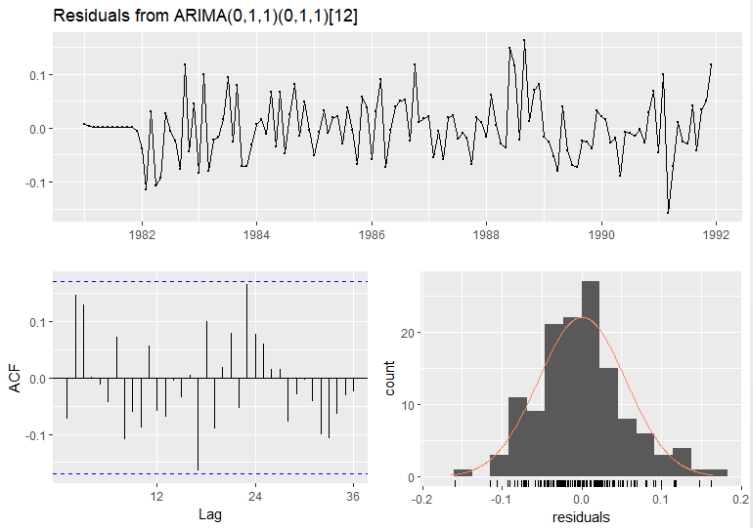
1) (p=0, q=1)(Q=1), (p=0, q=1)(P=1)에 의한 모형 탐색

- 모형 추정(p=0, q=1)(Q=1): ARIMA(0,1,1)(0,1,1)$_{12}$

```
> fit1 <- Arima(lntour,order=c(0,1,1),
                seasonal=list(order=c(0,1,1),period=12))
> fit1
Series: lntour
ARIMA(0,1,1)(0,1,1)[12]

Coefficients:
          ma1     sma1
      -0.5800  -0.5525
s.e.   0.0698   0.0791

sigma^2 estimated as 0.003351:  log likelihood=169.06
AIC=-332.13   AICc=-331.92   BIC=-323.79
```

- 모형 검진: ARIMA(0,1,1)(0,1,1)$_{12}$

```
> checkresiduals(fit1)

data:  Residuals from ARIMA(0,1,1)(0,1,1)[12]
Q* = 26.06, df = 22, p-value = 0.2491
```



- 과대적합: ARIMA(0,1,1)(0,1,1)$_{12}$

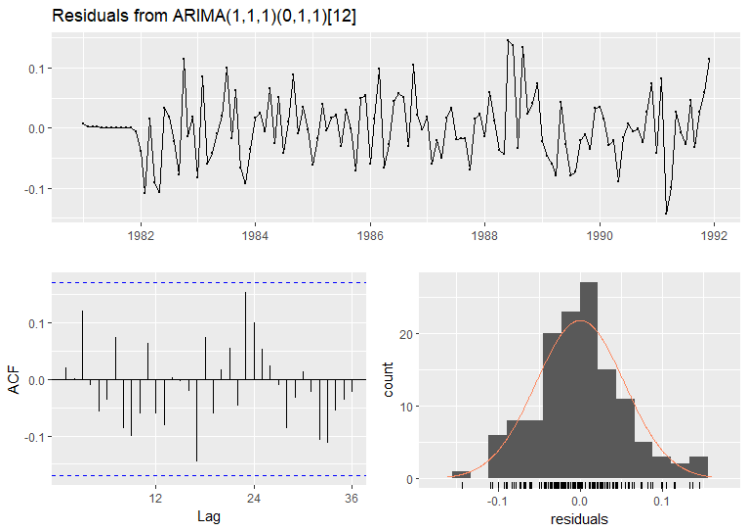
```
> fit1_1 <- Arima(lntour,order=c(1,1,1),
                  seasonal=list(order=c(0,1,1),period=12))
> fit1_2 <- Arima(lntour,order=c(0,1,2),
                  seasonal=list(order=c(0,1,1),period=12))
> confint(fit1_1)
           2.5 %      97.5 %
ar1  -0.6141189 -0.07646367
ma1  -0.6019970 -0.07213300
sma1 -0.6757900 -0.36348852
> confint(fit1_2)
            2.5 %     97.5 %
ma1  -0.84863210 -0.5006825
ma2   0.04584469  0.4197128
sma1 -0.67412800 -0.3580535
```

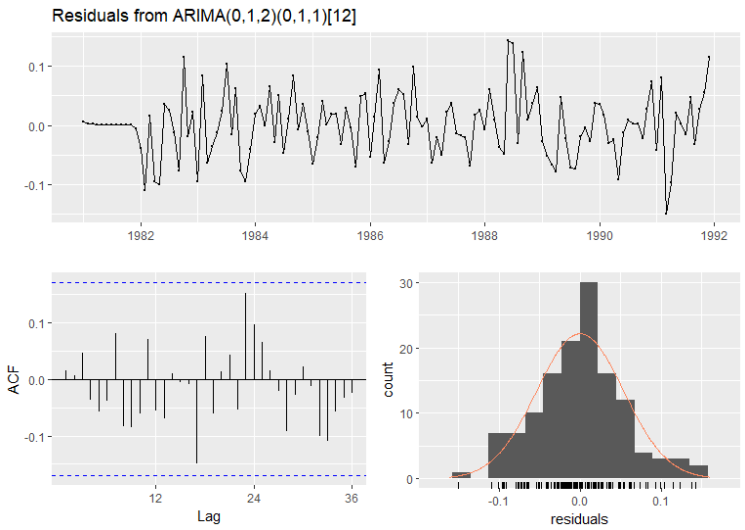
- 추가된 모수 유의적

- 과대적합은 비계절형 모수에만 적용

- 모형 검진: ARIMA(1,1,1)(0,1,1)$_{12}$

ARIMA(0,1,2)(0,1,1)$_{12}$






```
        Ljung-Box test

data:  Residuals from ARIMA(1,1,1)(0,1,1)[12]
Q* = 19.807, df = 21, p-value = 0.5335
```

```
data:  Residuals from ARIMA(0,1,2)(0,1,1)[12]
Q* = 17.468, df = 21, p-value = 0.6824
```

- 과대적합: ARIMA(1,1,1)(0,1,1)$_{12}$와 ARIMA(0,1,2)(0,1,1)$_{12}$

```
> confint(Arima(lntour,order=c(1,1,2),
                seasonal=list(order=c(0,1,1),period=12)))
          2.5 %      97.5 %
ar1  -0.7576853  0.48124313
ma1  -1.1513542  0.05538992
ma2  -0.2314544  0.55715140
sma1 -0.6724509 -0.35743032
> confint(Arima(lntour,order=c(2,1,1),
                seasonal=list(order=c(0,1,1),period=12)))
          2.5 %      97.5 %
ar1  -1.4005757  0.05563757
ar2  -0.6505574  0.21547715
ma1  -0.7633223  0.73372547
sma1 -0.6770033 -0.36266535
> confint(Arima(lntour,order=c(0,1,3),
                seasonal=list(order=c(0,1,1),period=12)))
           2.5 %     97.5 %
ma1  -0.88113628 -0.5041239
ma2   0.03417003  0.5055626
ma3  -0.24727265  0.1398392
sma1 -0.67243530 -0.3582209
```

추가된 모수 비유의적

→ ARIMA(1,1,1)(0,1,1)$_{12}$와 ARIMA(0,1,2)(0,1,1)$_{12}$: 예측에 사용 가능한 모형

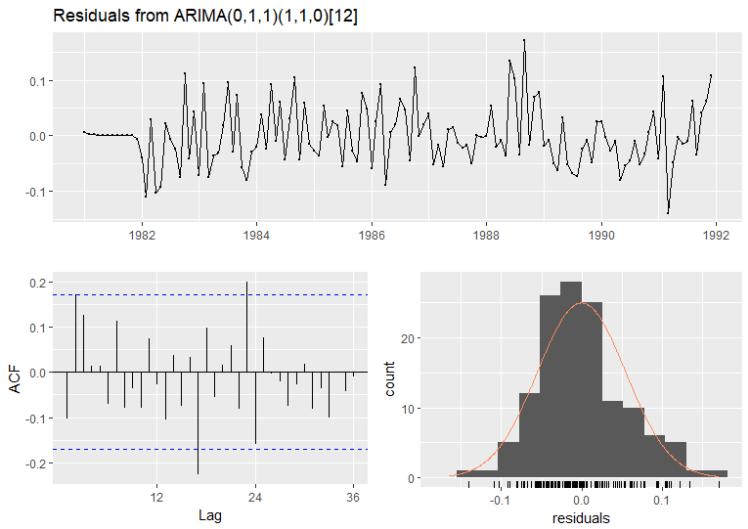
- 모형 추정 (p=0, q=1)(P=1) : ARIMA(0,1,1)$(1,1,0)_{12}$

```
> fit2 <- Arima(lntour,order=c(0,1,1),
                seasonal=list(order=c(1,1,0),period=12))
> fit2
Series: lntour
ARIMA(0,1,1)(1,1,0)[12]

Coefficients:
          ma1     sar1
      -0.6054  -0.5366
s.e.   0.0665   0.0810

sigma^2 estimated as 0.003346:  log likelihood=169.28
AIC=-332.56   AICc=-332.35   BIC=-324.23
```

ARIMA(0,1,1)$(0,1,1)_{12}$ 모형과 비슷한 예측력을 보임

- 모형 검진: ARIMA(0,1,1)(1,1,0)$_{12}$

```
> checkresiduals(fit2)

data:  Residuals from ARIMA(0,1,1)(1,1,0)[12]
Q* = 38.217, df = 22, p-value = 0.01732
```



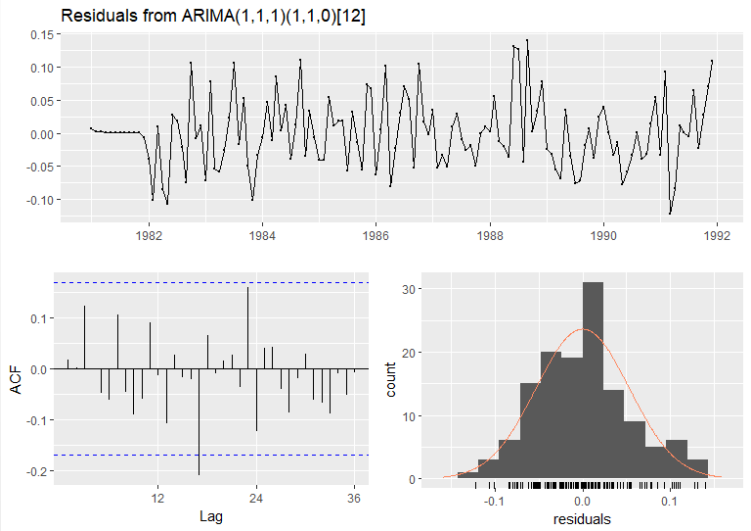
백색잡음 오차 가정 위반

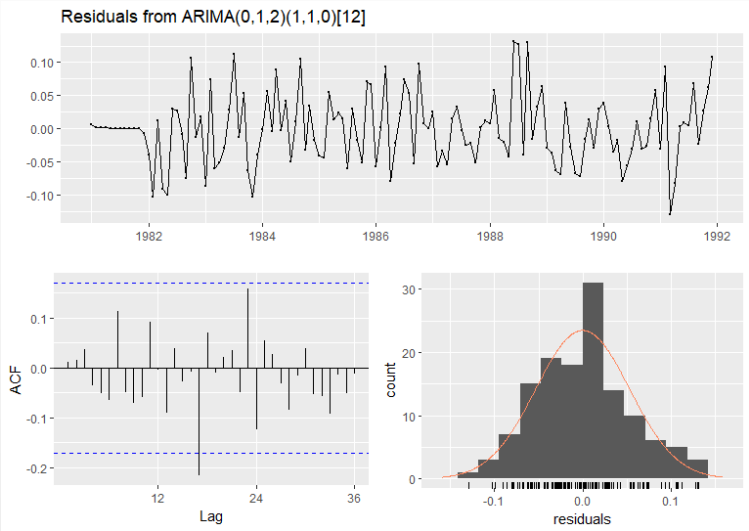
- 과대적합: ARIMA(0,1,1)(1,1,0)$_{12}$

```
> fit2_1 <- Arima(lntour,order=c(1,1,1),
                  seasonal=list(order=c(1,1,0),period=12))
> fit2_2 <- Arima(lntour,order=c(0,1,2),
                  seasonal=list(order=c(1,1,0),period=12))

> confint(fit2_1)
           2.5 %     97.5 %
ar1  -0.6244618 -0.1257088
ma1  -0.6016499 -0.1026214
sar1 -0.6862031 -0.3682476
> confint(fit2_2)
            2.5 %     97.5 %
ma1  -0.88940432 -0.5453766
ma2   0.08103535  0.4425904
sar1 -0.68381121 -0.3633908
```

추가된 모수 유의적

- 모형 검진: ARIMA(1,1,1)(1,1,0)$_{12}$ ARIMA(0,1,2)(1,1,0)$_{12}$






```
        Ljung-Box test

data:  Residuals from ARIMA(1,1,1)(1,1,0)[12]
Q* = 24.433, df = 21, p-value = 0.2726
```

```
data:  Residuals from ARIMA(0,1,2)(1,1,0)[12]
Q* = 22.927, df = 21, p-value = 0.3479
```

- 과대적합: ARIMA(1,1,1)(1,1,0)$_{12}$와 ARIMA(0,1,2)(1,1,0)$_{12}$

```
> confint(Arima(lntour,order=c(1,1,2),
                seasonal=list(order=c(1,1,0),period=12)))
          2.5 %     97.5 %
ar1  -0.7327746  0.4113774
ma1  -1.1291821 -0.0140854
ma2  -0.2036577  0.5600746
sar1 -0.6839650 -0.3642105
> confint(Arima(lntour,order=c(2,1,1),
                seasonal=list(order=c(1,1,0),period=12)))
          2.5 %       97.5 %
ar1  -1.3501035  0.003178749
ar2  -0.6346756  0.214711438
ma1  -0.7557297  0.634389249
sar1 -0.6863941 -0.367535856
> confint(Arima(lntour,order=c(0,1,3),
                seasonal=list(order=c(1,1,0),period=12)))
           2.5 %     97.5 %
ma1  -0.92461468 -0.5514994
ma2   0.07239749  0.5423576
ma3  -0.25181127  0.1276558
sar1 -0.68456703 -0.3650019
```

추가된 모수 비유의적

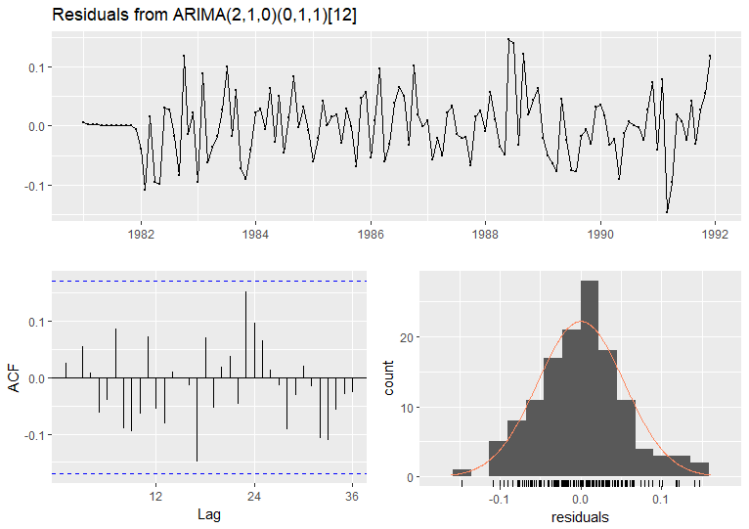
→ ARIMA(1,1,1)(1,1,0)$_{12}$와 ARIMA(0,1,2)(1,1,0)$_{12}$: 예측에 사용 가능한 모형

2) (p=2, q=0)(Q=1), (p=2, q=0)(P=1)에 의한 모형 탐색

- 모형 추정 (p=2, q=0)(Q=1): ARIMA(2,1,0)$(0,1,1)_{12}$

```
> fit3 <- Arima(lntour,order=c(2,1,0),
              seasonal=list(order=c(0,1,1),period=12))
> fit3
Series: lntour
ARIMA(2,1,0)(0,1,1)[12]

Coefficients:
          ar1      ar2     sma1
      -0.6864  -0.2253  -0.5199
s.e.   0.0908   0.0900   0.0802

sigma^2 estimated as 0.003237:  log likelihood=171.91
AIC=-335.81   AICc=-335.46   BIC=-324.7
```

- 모형 검진: ARIMA(2,1,0)(0,1,1)$_{12}$




```
        Ljung-Box test

data:  Residuals from ARIMA(2,1,0)(0,1,1)[12]
Q* = 18.225, df = 21, p-value = 0.6348
```

- 과대적합: 잠정모형 ARIMA(2,1,0)(0,1,1)$_{12}$

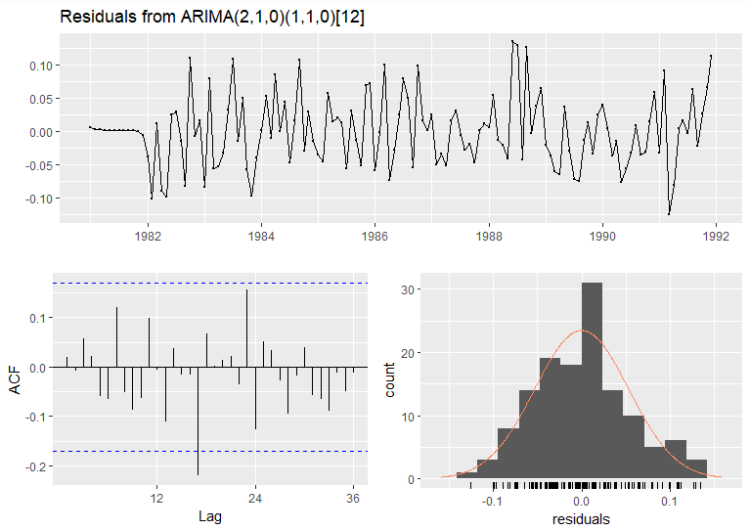
```
> confint(Arima(lntour,order=c(2,1,1),
                seasonal=list(order=c(0,1,1),period=12)))
           2.5 %      97.5 %
ar1  -1.4005757  0.05563757
ar2  -0.6505574  0.21547715
ma1  -0.7633223  0.73372547
sma1 -0.6770033 -0.36266535
> confint(Arima(lntour,order=c(3,1,0),
                seasonal=list(order=c(0,1,1),period=12)))
           2.5 %      97.5 %
ar1  -0.8719017 -0.50296810
ar2  -0.4487084 -0.00763568
ar3  -0.1888452  0.18083322
sma1 -0.6769962 -0.36266169
```

추가된 모수 비유의적 → 잠정모형을 예측모형으로 사용 가능

- 모형 추정 (p=2, q=0)(P=1) : ARIMA(2,1,0)(1,1,0)$_{12}$

```
> fit4 <- Arima(lntour,order=c(2,1,0),
                seasonal=list(order=c(1,1,0),period=12))
> fit4
Series: lntour
ARIMA(2,1,0)(1,1,0)[12]

Coefficients:
          ar1      ar2     sar1
      -0.7299  -0.2428  -0.5267
s.e.   0.0897   0.0894   0.0814

sigma^2 estimated as 0.003172:  log likelihood=173.04
AIC=-338.08   AICc=-337.73   BIC=-326.96
```

- 모형 검진: ARIMA(2,1,0)(1,1,0)$_{12}$




```
        Ljung-Box test

data:  Residuals from ARIMA(2,1,0)(1,1,0)[12]
Q* = 24.531, df = 21, p-value = 0.268
```

- 과대적합: 잠정모형 ARIMA(2,1,0)(1,1,0)$_{12}$

```
> confint(Arima(lntour,order=c(2,1,1),
                seasonal=list(order=c(1,1,0),period=12)))
          2.5 %       97.5 %
ar1  -1.3501035  0.003178749
ar2  -0.6346756  0.214711438
ma1  -0.7557297  0.634389249
sar1 -0.6863941 -0.367535856
> confint(Arima(lntour,order=c(3,1,0),
                seasonal=list(order=c(1,1,0),period=12)))
          2.5 %      97.5 %
ar1  -0.9181728 -0.55167096
ar2  -0.4827245 -0.03043428
ar3  -0.2034879  0.16766886
sar1 -0.6863178 -0.36759502
```

추가된 모수 비유의적 → 잠정모형을 예측모형으로 사용 가능

- AIC & BIC에 의한 최종 모형 선택

`fit1_1`: ARIMA$(1,1,1)(0,1,1)_{12}$ `fit1_2`: ARIMA$(0,1,2)(0,1,1)_{12}$

`fit2_1`: ARIMA$(1,1,1)(1,1,0)_{12}$ `fit2_2`: ARIMA$(0,1,2)(1,1,0)_{12}$

`fit3`: ARIMA$(2,1,0)(0,1,1)_{12}$ `fit4`: ARIMA$(2,1,0)(1,1,0)_{12}$

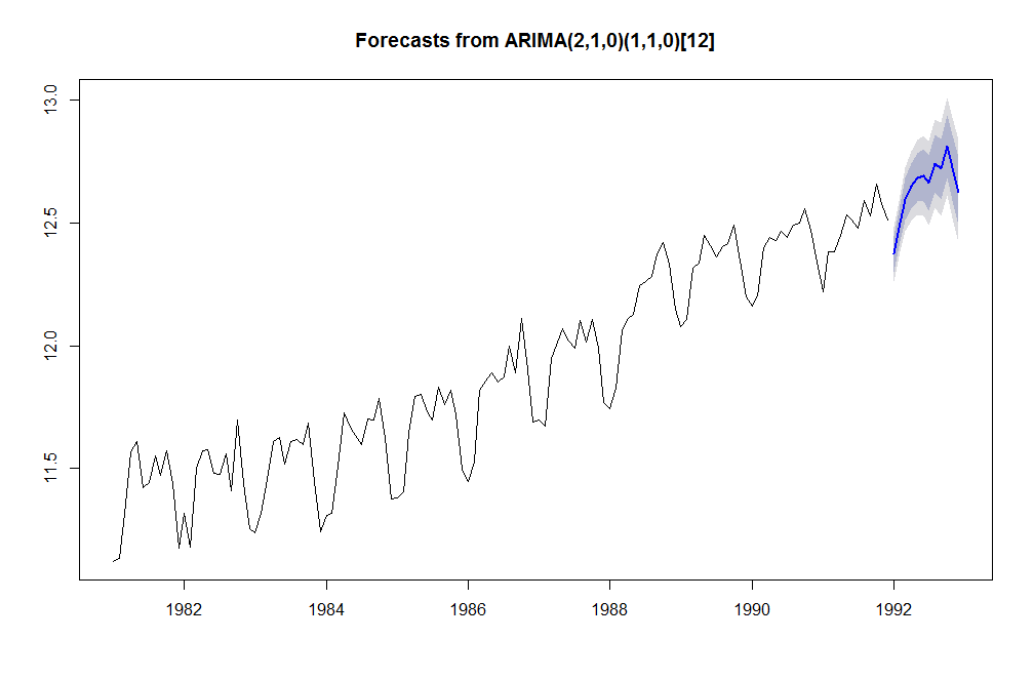
```
> c(fit1_1$aic, fit1_1$bic)
[1] -335.1006 -323.9841
> c(fit1_2$aic, fit1_2$bic)
[1] -335.4349 -324.3184
> c(fit2_1$aic, fit2_1$bic)
[1] -337.3768 -326.2603
> c(fit2_2$aic, fit2_2$bic)
[1] -337.7425 -326.6260
> c(fit3$aic, fit3$bic)
[1] -335.8118 -324.6953
> c(fit4$aic, fit4$bic)
[1] -338.0812 -326.9647
```

최종모형 : ARIMA$(2,1,0)(1,1,0)_{12}$

- 예측: ARIMA(2,1,0)(1,1,0)$_{12}$

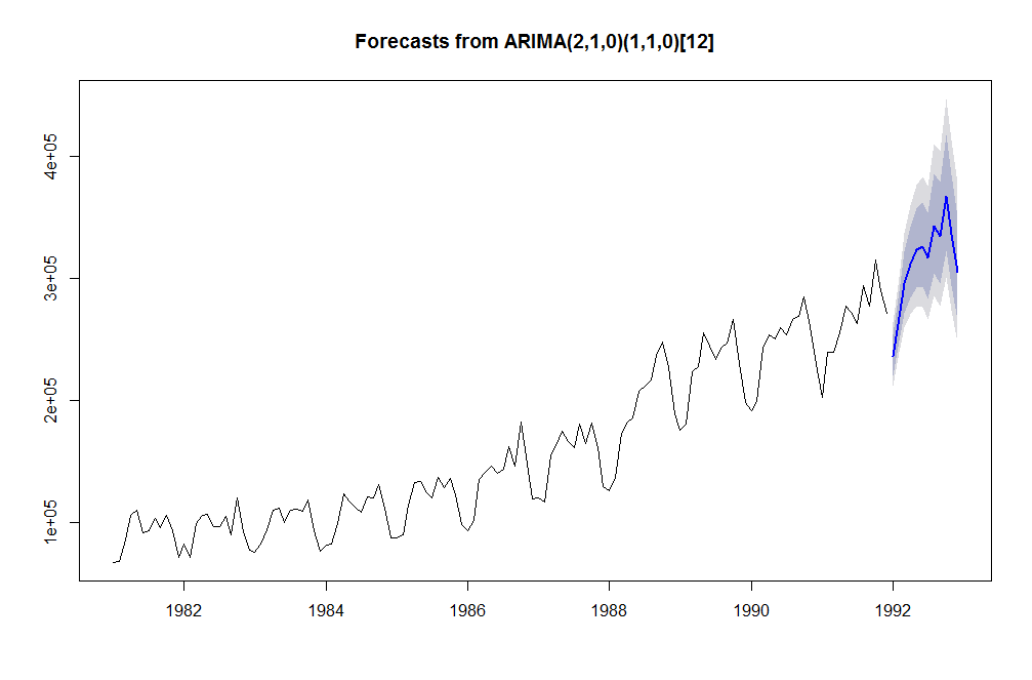
1) 로그 변환된 자료에 대한 예측

```
> fit4 <- Arima(lntour,order=c(2,1,0),
                seasonal=list(order=c(1,1,0),period=12))
> plot(forecast(fit4,h=12))
```



2) 원 자료(tour.ts)에 대한 예측

```
> fit4_1 <- Arima(tour.ts,order=c(2,1,0),
                  seasonal=list(order=c(1,1,0),period=12),lambda=0)
> plot(forecast(fit4_1,h=12))
```



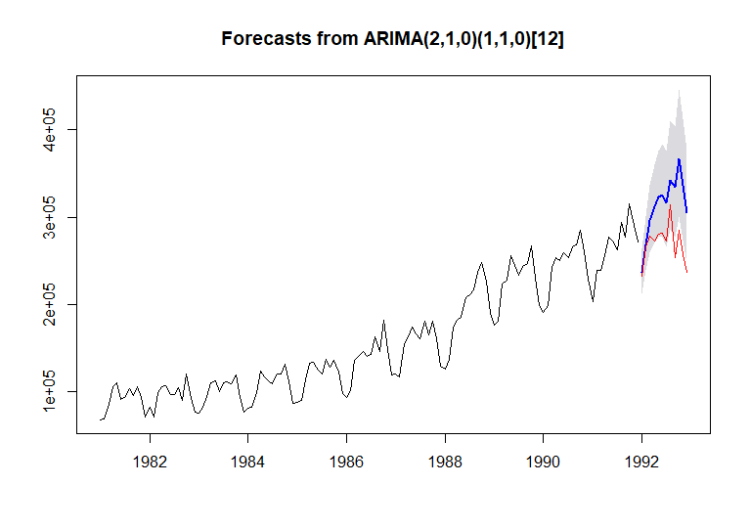
- 예측 결과와 실제 자료의 비교

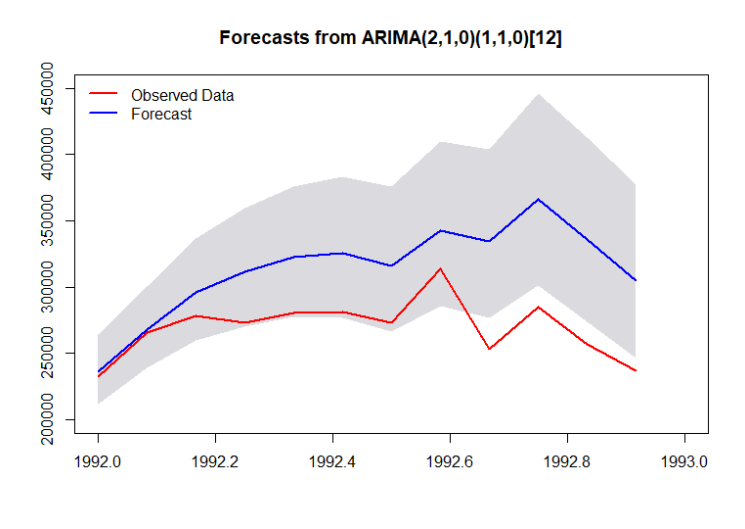
- 실제 자료의 입력

```
> tour92 <- scan("D:/Data/tour92.txt")
> tour92 <- ts(tour92,start=1992,freq=12)
```

- 예측 결과와 실제 자료의 시계열 그림 작성

```
> fore_arima <- forecast(fit4_1,h=12,level=95)
> plot(fore_arima)
> new_t <- seq(1992,by=1/12,length=12)
> lines(new_t,tour92,col="red")
```





```
> plot(fore_arima,xlim=c(1992,1993),ylim=c(2e+5,4.5e+5))
> lines(new_t,tour92,col="red",lwd=2)
> legend("topleft",c("Observed Data","Forecast"),lwd=2,
         col=c("red","blue"),bty='n')
```

- 예측 정확성 측도

```
> accuracy(fore_arima,tour92)
                      ME      RMSE       MAE          MPE     MAPE      MASE
Training set   -101.7388  9382.833  6691.828  -0.1666852  4.06514 0.3697756
Test set     -44218.3189 51964.159 44218.319 -16.5746164 16.57462 2.4434060
                  ACF1 Theil's U
Training set 0.1129040        NA
Test set     0.6654927  1.920409
```